R E P O R T R E S U M E S

ED 015 163 SP 001 396

EDUCATIONAL ATTAINMENT AND ATTITUDES TOWARD SCHOOL AS A FUNCTION OF FEEDBACK IN THE FORM OF TEACHERS' WRITTEN COMMENTS.

BY- SWEET, ROGER C.
WISCONSIN UNIV., MADISON

REPORT NUMBER TECHNICAL-REPORT-NO-15 PUB DATE NOV 66
REPORT NUMBER CENTER-NO-C-03
CONTRACT OEC-5-10-154
EDRS PRICE MF-$0.25 HC-$0.84 19P.

DESCRIPTORS- *ACADEMIC ACHIEVEMENT, BIBLIOGRAPHIES, *CHANGE AGENTS, *ENGLISH, FEEDBACK, *STUDENT ATTITUDES, STUDENT TESTING, STUDENTS, TABLES (DATA), TEACHING METHODS, *TEACHING TECHNIQUES, WILCOXON MATCHED-PAIRS SIGNED-RANKS TEST, FRIEDMAN 2 WAY ANOVA, OSGOOD'S SEMANTIC DIFFERENTIAL (EVALUATIVE DIMENSION)

RELATIONSHIPS BETWEEN A TEACHER'S WRITTEN COMMENTS ON A TEST AND (1) SUBSEQUENT STUDENT ATTAINMENT AS MEASURED BY TEST PERFORMANCE AND (2) ATTITUDE CHANGE TOWARD 9TH-GRADE ENGLISH, WERE INVESTIGATED. 9 CLASSES (225 STUDENTS) WERE GIVEN 4 TESTS BY THEIR TEACHERS OVER A 6-WEEK PERIOD. TESTS WERE RETURNED WITH NUMERICAL SCORES, LETTER GRADE, AND EITHER NO COMMENT (N), FREE COMMENT (F - WHATEVER TEACHER CARED TO MAKE), OR SPECIFIED COMMENT (S - DESIGNATED IN ADVANCE FOR EACH LETTER GRADE). ATTITUDE INVENTORY SCORES COLLECTED BEFORE THE FIRST AND AFTER RETURN OF THE LAST TEST WERE RANKED. ANALYSIS OF VARIANCE SHOWED THAT (A) F OR S COMMENTS HAD LITTLE IF ANY SHORT-TERM EFFECT ON TEST PERFORMANCE. (B) OVER A LONGER PERIOD OF TIME, F COMMENTS SIGNIFICANTLY INCREASED SCHOLASTIC PERFORMANCE, AND SIGNIFICANTLY CHANGED ATTITUDES IN A POSITIVE DIRECTION. INCLUSION OF S COMMENTS WAS NO MORE EFFECTIVE IN CHANGING ATTITUDES THAN WERE N COMMENTS. (AF)

BC01396

EDUCATIONAL ATTAINMENT AND ATTITUDES TOWARD SCHOOL AS A FUNCTION OF FEEDBACK IN THE FORM OF TEACHERS' WRITTEN COMMENTS

ED015163

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

RESEARCH AND DEVELOPMENT
CENTER FOR LEARNING
AND RE-EDUCATION

R600

DEC 26 1967

UNIVERSITAS WISCONSINENSIS SIGILLUM

RESEARCH
CLR
DEVELOPMENT

Technical Report No. 15

EDUCATIONAL ATTAINMENT AND ATTITUDES TOWARD SCHOOL AS A FUNCTION OF FEEDBACK IN THE FORM OF TEACHERS' WRITTEN COMMENTS

Roger C. Sweet

ed on a master's thesis under the direction of

Herbert J. Klausmeier,

Professor of Educational Psychology

Research and Development Center
for Learning and Re-Education
The University of Wisconsin
Madison, Wisconsin

November 1966

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education, and Welfare, under the provisions of the Cooperative Research Program.

Center No. C–03 / Contract OE 5–10–154

OTHER REPORTS OF THE R & D CENTER FOR LEARNING & RE-EDUCATION

TECHNICAL REPORTS

No. 1 Klausmeier, H. J., Davis, J. K., Ramsay, J. G., Fredrick, W. C., and Davies, Mary H. Concept learning and problem solving: A bibliography, 1950 - 1964. October 1965.

No. 2 Goodwin, W. L. The effects on achievement test results of varying conditions of experimental atmosphere, notice of test, test administration, and test scoring. November 1965.

No. 3 Fredrick, W. C. The effects of instructions, concept complexity, method of presentation, and order of concepts upon a concept attainment task. November 1965.

No. 4 Ramsay, J. G. The attainment of concepts from figural and verbal instances, by individuals and pairs. January 1966.

No. 5 Van Engen, H., and Steffe, L. P. First grade children's concept of addition of natural numbers. February 1966.

No. 6 Lynch, D. O. Concept identification as a function of instructions, labels, sequence, concept type, and test item type. June 1966.

No. 7 Biaggio, Angela M. B. Relative predictability of freshman grade-point averages from SAT scores in Negro and white Southern colleges. September 1966.

No. 8 Kalish, Patricia W. Concept attainment as a function of monetary incentives, competition, and instructions. September 1966.

No. 9 Baldwin, Thelma, L., and Johnson, T. J. Teacher behaviors and effectiveness of reinforcement. September 1966.

No. 10 Fang, M. C. S. Effect of incentive and complexity on performance of students from two social class backgrounds on a concept identification task. September 1966.

No. 11 Lemke, E. A., Klausmeier, H. J., and Harris, C. W. The relationship of selected cognitive abilities to concept attainment and information processing. October 1966.

No. 12 Burmeister, Lou E. An evaluation of the inductive and deductive group approaches to teaching selected word analysis generalizations to disabled readers in eighth and ninth grade. October 1966.

No. 13 Boe, Barbara L. The ability of secondary school pupils to perceive the plane sections of selected solid figures. November 1966.

OCCASIONAL PAPERS

No. 1 Staats, A. W. Emotions and images in language: A learning analysis of their acquisition and function. June 1966.

No. 2 Davis, G. A. The current status of research and theory in human problem solving. June 1966.

No. 3 Klausmeier, H. J., Goodwin, W. L., Prash, J., and Goodson, M. R. Project MODELS: Maximizing opportunities for development and experimentation in learning in the schools. July 1966.

No. 4 Otto, W. The relationship of reactive inhibition and school achievement: Theory, research, and implications. September 1966.

No. 5 Baker, F. B. The development of a computer model of the concept-attainment process: A preliminary report. October 1966.

PREFACE

This technical report is based on the master's thesis of Roger C. Sweet. Thesis committee members were Herbert J. Klausmeier, Chairman; Gary A. Davis; and Thomas Ringness.

In our program of research and development at the R & D Center for Learning and Re-Education, we have identified sets of variables related to five main categories -- stimulus material, instructions, response modes, conditions of learning, and organismic. The complete taxonomy is presented in Technical Report No. 1 of the Center.

In the present study, Mr. Sweet examined the relationship between the teacher's written comments on a test paper (informative feedback) and subsequent attainment and attitudes toward a school subject, namely, ninth grade English.

Herbert J. Klausmeier
Co-Director for Research
Professor of Educational Psychology

CONTENTS

		Page
	List of Tables	vii
	Abstract	ix
I.	Introduction	1
II.	Procedure	4
	Subjects	4
	Data Gathered	4
	Method	4
	General	4
	Scholastic Achievement	5
	Attitude Change	6
	Statistical Treatment	6
III.	Results	7
IV.	Discussion	9
	Scholastic Achievement	9
	Attitude Change	10
	Concluding Statements	10
	References	11

LIST OF TABLES

Table		Page
1	Illustration of Ranked Data	5
2	Illustration of the Number of Levels in Each Class	5
3	Procedure Used for Reranking Data	6
4	Friedman Test of Over-all Treatment Effects	7
5	Friedman Analysis of Reranked Data (K = 2) Across Individual Subjects	7
6	Attitude Change Based on the Wilcoxon Matched-Pairs Signed-Ranks Test	8

ABSTRACT

The study dealt with a partially established relationship between a teacher's written comments on a test and subsequent student attainment as measured by test performance. In addition, an attempt was made to ascertain any attitude change towards a particular school subject (9th grade English) as a function of the teacher's comment. Three classes of each of three teachers comprised the sample of 225 students taking ninth grade English. Over a period of six weeks, each teacher gave four tests which were not of the long essay type. When each test had been corrected, the teacher returned the test papers with the numerical score and letter grade as earned. No Comment (N) students received only the numerical score and letter grade. Free Comment (F) students received whatever comment the teacher felt it desireable to make. The Specified Comment (S) students received comments designated in advance for each letter grade. Attitude inventory scores, based on Osgood's semantic differential (evaluative dimension), were collected on the day before the administration of the first test and soon after the return of the fourth test. Because of the qualitative and quantitative differences between all the different tests used by different teachers, the tests were regarded as ranking instruments.

The Friedman Two-Way ANOVA was used to analyze the ranked data across individual subjects and across classes. Comments of either a free or specified nature have little if any short-term effect on test performance; over a longer period of time, the inclusion of free comments has a significant effect on scholastic performance. Attitudes were analyzed by using the Wilcoxon Matched-Pairs Signed-Ranks test. A highly negative Z signified that only under the free comment condition were attitudes significantly changed in a positive direction. This indicated that the inclusion of specified comments was no more effective in changing attitudes than were no comments.

I

INTRODUCTION

The following investigation dealt with a partially established relationship between a teacher's written comments on a test and subsequent student attainment as measured by test performance (Page, 1958a). In addition, an attempt was made to ascertain any attitude change towards a particular school subject (9th grade English) as a function of the teacher's comment.

If teachers' comments were treated as social reinforcement within the usual S-R framework, a multitude of theoretical questions would arise for which there were no readily available answers. Treating comments as a reinforcer leads us to the question of what might happen if the comments were discontinued. According to the basic principles of S-R theory, extinction should follow. However, in a high school classroom, it would be difficult indeed to decide just exactly what it is that is being extinguished. As Page (1958a, 1958b) pointed out, investigations of praise and blame have provided very fruitful knowledge for the general psychologist. For the educator, however, these same investigations are belabored by many weaknesses. It was this author's opinion that most experimental attempts to measure the effects of praise and blame have been accomplished under specially arranged situations where the effects of extraneous factors have been minimized and the verification of basic S-R principles virtually assured. This does not deny the great value of these basic principles to the area of human learning; however, it does suggest that the extension to the classroom may not provide an adequate test of the theory.

A teacher's comment may be more acceptably treated as a type of feedback. Further, as Bilodeau and Bilodeau (1961) stated, "Studies of feedback or knowledge of results show it to be the strongest most important variable controlling performance and learning." The comments of teachers in the present study could be viewed as feedback in that the student noted his errors and correct responses and saw the grade and the evaluative comments which were put on his test paper. One type of feedback was the error corrections. The student was told the direction and extent of his errors with the information supplied in the form of error corrections, thus serving the directing function of feedback.

Teachers' comments are also feedback, but feedback of a different nature. Like letter grades, the comments are a reinforcement component providing feedback to the student about some of the effects of his behavior. They are examples of the teacher's communication of approval or disapproval over the student's work. A look at the index of the Psychological Abstracts may aid in illustrating this point. Under the word "Feedback" is the statement, "See also Knowledge of Results and Reinforcement." Solomon and Rosenberg (1964) broke down feedback in a very similar manner. Their article was intended to illustrate how teacher-student feedback could affect the social structure of the classroom. Though their particular problem holds little immediate relevance to the present topic, their method of analyzing the concept of feedback is of great importance. They wrote of an informational as well as a reinforcement component. By an informational component, they meant indicating correct answers (telling the student that an answer is right, if it is; or telling him that it is wrong and providing the correct answer). The reinforcement component referred to the teacher's communication of approval or disapproval. This dualistic conception of feedback does not differ from the informational and reinforcement components involved in programmed instructional sequences. Schvaneveldt (1965), in an extremely thorough review of the informational component of feedback, stated that a performance-related signal may be called anything from reward to knowledge of performance, because of the dimensions on which it could vary.

Though it proves impossible to dismiss all the present ambiguity surrounding such

terms as feedback and reinforcement, the present writer chose to treat teachers' comments as being, for the most part, a reinforcement component of feedback, as opposed to an informational or knowledge of results component.

A close analysis of Page's (1958a) study reveals a very nontheoretical attitude. Page used the terms "praise" and "blame," thus implicitly considering teacher's comments to be reinforcers; however, throughout the paper, words such as "reinforcement," "extinction," and "secondary reinforcers" were conspicuously absent. It was believed that this very empirical and pragmatic philosophy would prove to be the most efficient way of handling the present problem.

A survey of the literature reveals a great deal of latitude with regard to the above philosophy. The studies range from the very nontheoretical article of Page's to the rather narrow S-R formulations of Skinner. However, a great deal of credit belongs to Skinner for the reason that he brought the popular misconception that "learning is its own best reward" to task (Skinner and Rogers, 1956) while criticizing the commonly held belief that the learning process, or knowledge itself, is attributable to something inside the individual. Furthermore, Skinner and Holland (1960) felt that compliance with the above attitudes automatically put the entire responsibility for learning upon the student, giving little regard to any possible inadequacies of the training program. They were of the opinion that the responsibility for learning should be carried by the teacher and the teaching situation. In his analysis of Skinnerian methodology, Hively (1959) also elaborated on who was to be responsible for learning in suggesting that the reinforcement function in the control of behavior implied a series of operations analogous to those employed by a skilled private tutor.

In a further elaboration of this idea, Staats and Staats (1962, 1963), discussed "achievement behaviors." They introduced the theory that, in a naive individual, overcoming obstacles and doing something difficult was not itself originally reinforcing. For some children, working at certain tasks such as school work may be heavily reinforced. The parents may be the source for much of this reinforcement, especially if they are seen as highly reinforcing by the child. If this is the case, any type of feedback paired with parental approval (good grades) should also take on reinforcing properties. Thus, a child who is raised in an environment where objects and events pertaining to school have been in contiguous association with positive reinforcers should find a more abundant supply of reinforcing stimuli in the school situation.

In an article specifically related to the effectiveness of verbal reinforcement, McDavid (1959) concluded that the more effective social approval is as a reward, the greater the motivational or incentive value, and consequently, the greater the probability of high scholastic achievement. Furness (1958) accomplished a very thorough analysis of all factors, both environmental and organismic, involved in successful spelling behavior and concluded that verbal reinforcement is one of the most important.

The effects of praise and blame as a function of intelligence have been investigated by Kennedy, Turner, and Lindner (1962). Relevant to the present investigation was the fact that they studied the effects of praise and blame without using formal S-R terminology. The one unfortunate aspect, however, revolved around their learning task which was a simple visual discrimination problem, far removed from the normal course work of high school students.

There has been a vast amount of research revolving around the effects of feedback, yet very few investigators have studied the effects of written comments on test papers. The most exhaustive study of this variable was accomplished by Page (1958a). Page used 74 randomly selected secondary teachers, who were teaching a total of 2,139 students. The teachers administered to their respective students whatever objective tests would occur in the usual course of instruction. After scoring and grading the test papers in their customary way, and matching the students by performance, they randomly assigned the papers to one of three treatment groups. The No Comment group received no marks beyond those for grading. The Free Comment group received whatever comments the teachers felt were appropriate for the particular students and test. The Specified Comment group received certain uniform comments, designated beforehand by Page for all similar letter grades, which were felt to be generally "encouraging." The teachers returned the tests to the students without any unusual attention. The scores on the next objective test became the criterion of comment effect. Page found that students who had received either a free or specified comment on the first test did significantly better on the second test than did those students who received nothing but a

numerical score and letter grade on the first test. These results held across different schools, ages, and grade-point averages of the students involved. The opinion held by many teachers that the better students would be more responsive to the comments was not verified.

Page's study demonstrates two very important points. First, it illustrates that methodologically "good" research can be done in a normal classroom setting. Thus, we find present one of those rarities of psychological research, a well controlled and well designed study whose data was of immediate use somewhere besides the laboratory. Second, Page provided proof that a combination of both informational (grades) and reinforcing (personalized comments) feedback is superior to informational feedback alone in positively affecting scholastic performance. Related to the idea of comment inclusion is the hypothesis that a student who sees a comment, positive or negative, on his work will feel that the teacher must really be interested in him. This inference was in agreement with the results of Prosad and Singh (1962), who illustrated that undergraduate students feel that the better teachers are the ones who show an individual interest in them.

In addition to analyzing the effects which written comment inclusion may have on scholastic performance, the present study intended to ascertain what effects, if any, these comments may have in changing a student's attitude. The popular conception that scholastic achievement and attitudes toward school are closely related has received a great deal of empirical support: Quay (1959), Bostrum, Vlandis, and Rosenbaum (1961), Brodie (1964), Aiken and Dreger (1961), and Weaver (1959). However, Wright and Jung (1959) presented 1,011 excellent reasons for not considering this relationship automatic. They investigated the reasons that 1,011 students who finished in the top 10% of their high school class did not desire to continue their education. Among the most often stated reasons was a specific dislike for school and associated factors.

Normally, attitudes and behavior have been considered as two separate entities, with one seen as causing a change in the other. The majority of studies in this area have dealt with the causal effects of behavior change upon the changing of an attitude. The social psychological research of Festinger centers around the theory of cognitive dissonance as an explanatory concept. His model, as well as other cognitive consistency models of attitude change ("Congruity," "Balance"), are reviewed by Cohen (1964).

For this study, the present writer was unable to see the necessity for treating behavior and attitude in a causal fashion. Rather, it was decided to treat these entities as explicit (behavior) and implicit (attitude) responses, which were subject to change not as a function of each other, but rather as a function of feedback in the form of teacher's comments.

The following hypotheses were suggested:

A. Returned test papers bearing Free and Specified written comments of teachers, along with letter or numerical grades and error corrections, are associated with higher student attainment in ninth grade English over a short time period, more so than are tests which are returned with No Comments, merely containing a letter or numerical grade and error corrections.
B. Over a longer period of time only Free written comments by teachers, along with letter or numerical grades and error corrections on returned test paper, are associated with higher student attainment.
C. Attitudes toward ninth grade English are positively influenced by the inclusion of Free Comments or Specified Comments but not by No Comments.

Indirectly, Hypothesis A is opposed to the findings of Lintz and Brackbill (1966) whose comparisons of money rewards and flashing lights have shown little effect on performance. Schvaneveldt (1965) also felt that experiments with human adults have demonstrated null effects regarding the manipulation of reward independent of information. While it is agreed that the inclusion of written comments without error corrections or letter grades would not be sufficient to improve school performance, it is also felt that a combination of written comments with error corrections and letter grades is superior in improving school attainment to a test paper being returned which contains only error corrections and a letter grade.

Hypothesis B represents an elaboration on the findings of Page (1958a) who found no significant differences in the effects of Free and Specified Comments on school performance. His study only covered the amount of time it took to administer two tests. It is suggested that as more tests are given over a greater period of time, the student will become "immune" to specified or stock comments.

II
PROCEDURE

SUBJECTS

The sample was drawn from students taking ninth grade English at a high school in a large Midwest city. From all teachers instructing ninth grade English at the school, three were chosen at random. From each of the three teachers, three classes were randomly chosen. The final N consisted of 225 students.

DATA GATHERED

During a period of six weeks, each teacher gave four tests which were not of the long essay type. (See Sweet, 1966, for a more detailed description of the instructions given to the teachers.) When each set of tests had been corrected, the teacher returned the test papers with the numerical score and letter grade as earned. Each test was returned before another one was given. The experimental treatments were as follows (Page, 1958): No Comment (N) students received no comments, just the numerical score and letter grade. Free Comment (F) students received whatever comment the teacher felt it desireable to make. The Specified Comment (S) students received comments designated in advance for each letter grade as follows:

A - Excellent! Keep it up.
B - Good work. Keep it up.
C - Perhaps try to do still better.
D - Let's bring this up.
F - Let's raise this grade!

Attitude inventory scores were collected on two occasions (See Sweet, 1966). The students filled out the inventory the day before the administration of the first test and soon after the return of the fourth test. It took about twenty minutes on each occasion to administer the inventory. The inventory developed for this study was based on Osgood's semantic differential (Osgood, Suci, and Tannenbaum, 1957) and used the three scales which were shown to be most factorially pure with regard to the "evaluative" dimension.

The decision to use the differential in this study was based on Osgood's opinion that an attitude is some portion of the internal mediational activity and thus a part of the semantic structure of the individual. This opinion originated from the idea of an attitude as being a learned implicit response which is potentially bipolar and which mediates all evaluative behavior. It was decided to use the differential in this study because of the above opinions, and because an apparent "evaluative dimension" had been isolated.

The following is a sample from the differential, and the "responses" to a particular item:

Charles Dickens

Good	1	2	3	4	(5)	6	7	Bad
Valuable	1	2	3	(4)	5	6	7	Worthless
Positive	1	2	(3)	4	5	6	7	Negative

For the item "Charles Dickens," the student would have a total score of twelve. For each item, the possible score ranged from three to twenty-one.

METHOD

General

The students were assigned to one of the three treatments in the following manner. The students of each teacher within each class were ranked according to their first semester grade in English. In this respect, the present study differed from that of Page, who ranked and assigned to a treatment group on the basis of the scores on the first objective

test. Each of the top three students of each teacher was randomly assigned to one of the three treatment groups. This procedure was then repeated with the next best three, and so on, until all students had been assigned to one of the treatment groups. In all, there were 225 students, making a total of 75 subjects within each treatment.

Scholastic Achievement

The effects of comments were judged by the scores achieved on the second and fourth test, regardless of the nature of that test. This was done to investigate the "short term" and "long term" effects of the comments and thus verify or refute Hypothesis B. Certain statistical problems were present, since each test differed from the others with regard to practically every conceivable test variable such as subject matter, length, and difficulty. However, when the tasks were regarded primarily as ranking instruments, most of the difficulties disappeared. For example, a class of twenty-seven students formed nine levels on the basis of the first semester grades in English. Each level consisted of three students, with each student receiving a different treatment: No Comment, Free Comment, or Specified Comment. Students then were given raw scores on each of the four tests within a six-week period. On the basis of such scores, they were assigned rankings within levels as illustrated in Table 1, Part B.

Table 1

Illustration of Ranked Data

	Part A (Raw Scores on Test 2) Treatment			Part B (Ranks within levels on Test 2) Treatment		
Level	N	S	F	N	S	F
1	33	31	34	2	1	3
2	30	25	32	2	1	3
3	29	33	23	2	3	1
.	..	..	..	.	.	.
9	14	25	21	1	3	2
Sum				19	21	20
Sum Ranks				1	3	2

Note: Taken From Page (1958a), p. 312.

In all, there were nine classes, consisting of 75 levels making up a total 225 students. This is illustrated in Table 2.

Table 2

Illustration of the Number of Levels in Each Class

Class	No. of Levels per Class	
1	7	(21 Students)
2	8	(24 Students)
3	9	(27 Students)
4	9	(27 Students)
5	9	(27 Students)
6	9	(27 Students)
Total	75	225

The Friedman Two-Way Analysis of Variance was used to analyze the treatment effects when students were considered as matched independently from one common population. In this case, the summation of the rankings of 75 levels were analyzed. In addition, the Friedman was used analyze the treatment effects when treatment groups within classes were regarded as intact groups. For this particular analysis, the sums of the the ranks in each class were ranked. Referring back to the bottom of Table 1, the sums 19, 21, and 20 are ranked. Their rankings would be 1, 3, and 2, respectively. This procedure was carried out nine times, once for each of the nine classes.

In order to analyze the effects of only two treatments at a time, one treatment was dropped out, and the other two were reranked. In this situation, the number of treatments (K) was equal to two, which introduced the problem of the feasibility of using the Friedman in such an analysis. Friedman (1937) stated that in this special case (K = 2) the method of ranks was equivalent to the binomial series test, which is equivalent to the sign test when $N > 25$. Class-group data could not be computed in this fashion because of the fact that the number of the distribution no longer approaches normality when $N < 25$. (N = 9).

Table 3

Procedure Used For Reranking Data

Part A Raw Score			Part B Ranking (N, S, F)			Part C ReRanking		Part D ReRanking		Part E ReRanking	
N	S	F	N	S	F	F	S	F	N	S	N
33	31	34	2	1	3	2	1	2	1	1	2

Note: For a more complete illustration, see Appendix to Sweet (1966).

Attitude Change

Attitudes toward ninth grade English were measured using the three most factorially pure scales of Osgood's evaluative dimension. Scoring was arranged so that smaller scores indicated more positive attitudes. The Wilcoxon Matched-Pairs Signed-Ranks test was used in order to discover any possible attitude change. It was far more powerful than the simple sign test in that it utilized information about magnitude, as well as direction of change. It gave more weight to a pair which showed a large difference between two conditions, than to a pair which showed a small difference.

STATISTICAL TREATMENT

A nonparametric test, the Friedman two-way ANOVA, was used to compare the effects of teacher's comments on test performance. This test is useful when the measurement of the variable is in at least an ordinal scale. It was previously stated (page 6), that the Friedman was used in analyzing the overall treatment effects both across all students, and between the nine classes. This was accomplished in the following manner:

(a) Across individual subjects. In this case, the summations of the rankings for seventy-five levels were analyzed. The three grand sums (one for each treatment), were then used to compute the value of Xr^2 using formula 2a (In Siegel, 1956):

$$Xr^2 = \frac{12}{NK(K+1)}\left[(R_1)^2 + (R_2)^2 \ldots + (R_j)^2\right] - 3N(K+1),$$

where K = number of treatments (3)
N = number of levels or rows (75)
R = sum of the ranks for each column

(b) Across classes. Here, the Friedman (formula 2a) was used to compare the treatment effects when treatment groups within classes were regarded as intact groups. In this situation, N = 9 rather than 75.

Formula 2a was also applicable in analyzing only two treatments at a time. Here, K = 2, not 3. This analysis could not be undertaken for comparing between class effects for reasons already stated on page 6.

The Wilcoxon test necessitated the computation of T, the statistic on which the Wilcoxon is based. In order to compute T, let d equal the difference between the score on the first and second administration of the inventory. All of the d's were then ranked without regard to sign. Then, the sign of the difference was affixed to each rank, indicating which ranks arose from negative d's (positive change in attitude), and which ranks arose from positive d's (negative change in attitude). The number of ranks having a + d and a - d were tabulated. The Wilcoxon T is the summation of those ranks having the least frequent sign. T was the summation of positive ranks under all three treatments indicating that most students regardless of treatment, experienced a positive change in their attitudes toward ninth grade English.

Once the Wilcoxon T was computed, it was introduced into formula 2b (Siegel, 1956) which is as follows:

$$Z = \frac{T - M_T}{\sigma_T},$$

with M_T = Mean $= \frac{N(N+1)}{4}$

and σ_T = SD $= \sqrt{\frac{N(N+1)(2N+1)}{24}}$,

and N = Number of levels.

With N > 25, the sum of the ranks, T, is approximately normally distributed, allowing for the computation of a Z score (formula 2b).

III

RESULTS

Table 4 illustrates the overall comparison of all three treatment effects across individual subjects and classes for Tests 2 and 4. From Test 2 to Test 4, the performance of students under treatment N remained stable, while deteriorating under treatment S, and improving under treatment F. Test 2, or short term data, indicated little if any treatment effect with regard to individual subjects. Treated as independent class groups, however, some treatment effect was noticed with the probability of getting a Xr^2 of 4.832 or larger no greater than 10%.

The probability ($P < .15$) associated with the observed Xr^2 of 3.845 for the Test 4 individual-subjects data, indicated that there may be a definite trend with regard to the effects of written comments. Once again, there was a moderately low probability ($P < .10$) associated with the Test 4 class groups data.

Table 5 lists the results when only two treatments at a time were compared. The Test 2 data (short term) seemed to present little evidence favoring either Specified or Free Comments over No Comments. However, trends in this direction were present as witnessed by the relatively small probabilities associated with the observed Xr^2 values between FN and SN in comparison with the large probability ($P < .60$) associated with the observed Xr^2 values between treatments F and S.

The Test 4 data (long term) indicated that the majority of any treatment effect was related to the Free Comment condition.

Table 4

Friedman Test of Overall Treatment Effects

		N	S	F	df	Xr^2	P
Test 2	Individual Subjects	140.5	158	151.5	2	2.086	<.35
	Class-group Subjects	17.5	20.5	17	2	4.832	<.10
Test 4	Individual Subjects	140.5	146	163.5	2	3.845	<.15
	Class-group Subjects	13.5	18	22.5	2	4.50	<.10

Note: Modeled after Friedman in Siegel (1956), pp. 166-173.

Table 5

Friedman Analysis of Reranked Data (K = 2)
Across Individual Subjects

		N	S	F	df	Xr^2	P
	Between F and S		115	110	1	.33	<.60
Test 2	Between F and N	107		118	1	1.61	<.20
	Between S and N	106	119		1	2.25	<.15
	Between F and S		107	118	1	1.61	<.20
Test 4	Between F and N	103		122	1	4.81	< 03
	Between S and N	110	115		1	.33	< 60

This was evident in all three Test 4 comparisons. In the comparison between treatments F and S, the probabilities associated with the observed Xr^2's dropped from 60% (Test 2) to 20% (Test 4), while the probabilities associated with the Xr^2 values between S and N rose from 15% (Test 2) to 60% (Test 4). The comparison between treatment F and treatment N indicated that there was only a 3% chance of being wrong in considering these two ranked summations as being from different populations.

Table 6 seems to clearly illustrate that only under the Free Comment treatment were attitudes significantly changed in a positive direction. This effect was significant at the 5% level. The Specified Comment and No Comment conditions were almost equally nonsignificant, indicating that the inclusion of Specified Comments was no more effective in changing attitudes than were No Comments.

Table 6

Attitude Change Based On The Wilcoxon Matched-Pairs Signed-Ranks Test

Treatment	Wilcoxon T	Mean	Sd	N	Z	P
F	950.5	1387.5	185.62	74	-2.35	< .05
S	1238.5	1425	189.37	75	- .98	< .35
N	1155.5	1350.5	181.89	73	-1.07	< .30

Note: Modeled after Wilcoxon in Siegel (1956) pp. 75-83.
Whenever an individual received the same score on both administrations of the inventory, the score was dropped from the analysis, thus explaining why N does not equal 75 in the F and N treatments.

IV
DISCUSSION

SCHOLASTIC ACHIEVEMENT

Before considering the theoretical reasons associated with the results listed in Tables 4 and 5, there are certain statistical considerations which must be discussed. In analyzing overall treatment effects, Page (1958a) treated his ranked data in three ways. Two of these methods, which have already been discussed (Table 1), were also used in the present investigation. First, all the ranks were summed, making it possible to analyze the overall treatment effects with regard to individual subjects. Second, the sums of the ranks within each class were ranked. This allowed for the analysis of overall treatment effects with regard to class-group subjects. The Friedman Two-Way ANOVA is a nonparametric test lacking the power inherent in the parametric F test.

Page also analyzed his overall treatment effects in a third fashion. He took the summation of ranks in each column within each of his seventy-four classes, and then divided this sum by the number of levels in each class, with the result being a mean rank within treatment within class. In his reasons for doing this, Page commented, "This score proved very useful since it fulfilled certain requirements for parametric data [p. 313]."

Keeping in mind the statistical fact of life that a not very sensitive nonparametric test was used, the overall Test 2 results listed in Table 4 give limited support to the Page study and the predictions stated in Hypothesis A. Depending upon how large an one is willing to accept, the results listed in the Test 2 data of Table 5, could either support or dispute the Page findings and the short term prediction of treatment effects given in Hypothesis A. For instance, the short term results in Tables 4 and 5 could be interpreted in two ways. First, it could be argued that some promising trends were evident, that both the F and S treatments seemed to have the same short-term positive effect upon test performance. A second argument would be that by the end of the second test, any effects which comment inclusion, either free or specified, had were negligible. The first argument would lend support to the Page data, and the verification of Hypothesis A. However, the data in Table 3 cannot be ignored. While a possible trend may be present, the overall treatment effects with regard to individual subjects were very discouraging, thus causing the writer to question the Page results, being of the opinion that neither treatment had any appreciable short term effect on test performance.

A study of the long term (Test 4) data did indicate a definite treatment effect. The Table 4, Test 4 results provided the first hint of long term effects, across both individual subjects and classes. Table 5 indicated more specifically that any treatment effect was due mainly to the Free Comment condition. This would tend to support the predictions of Hypothesis B, but was definitely not in agreement with Page, who in his analysis after two tests found no significant differences between Free and Specified Comments.

It was hypothesized that Specified Comments would lose their effect by the time the fourth test was taken, the reasons being that the students would rapidly become "immune" to the appearance of "stock" comments. Psychophysical correlates of this notion come from "stimulus satiation" studies. Wolfle (1951) states that if training is of a prolonged and monotonous nature, variety in the stimulus materials may speed up learning rather than retard it. Seashore (1944) attempted to combat monotony in the training of radio operators by using a great variety of drill materials. Men trained under these conditions learned more rapidly than did men having less varied drills. Even though the stimulus materials being varied in the above two studies are not specifically related to feedback, the results are easily applicable to an explanation concerning the lack of long term treatment

effects with Specified Comments. It could be postulated that students become bored after seeing the same comment (stimulus) time and time again, and that something similar to the "satiation effects" found in lab studies could be occurring.

One further explanation concerning the lack of improvement under the specified condition revolves around the impersonal nature of the Specified Comments. It is possible that allowing the teacher to write what she desired on the student's paper caused her to make her comments far more personal, since she was probably familiar with the past progress and general makeup of each of her students.

No definite answer can yet be given for the effects, or the lack of them, which Specified Comments have, though the problem is interesting and easily open to investigation. At this point, all that can be said with certainty is that comments of either a free or specified nature have little if any short term effect on test performance, but that over a longer period of time the inclusion of Free Comments has a significant effect on scholastic performance.

ATTITUDE CHANGE

The most striking results came from the effects which comments had on the attitudes toward English. It was hypothesized that comments of either type would tend to change attitudes in a positive direction. The evidence presented in Table 6 did not entirely coincide with the predictions of Hypothesis C. Like the test performance data, it seemed that only Free Comment inclusion had a positive effect on attitude change. These results provided fairly good evidence supporting the opinion that attitudes are simply responses which are governed by many of the same laws related to other behaviors.

Of extreme interest was the fact that Specified Comments did not have any significant effects on attitude change. This set of data could serve to supplement the results concerning the authenticity surrounding the effects of Specified Comments on test performance. It illustrated that under certain conditions, there is a very strong R-R relationship between attitudes and test performance in ninth grade English. It could be hypothesized that we learn hierarchies of verbal and nonverbal responses, and that this is the reason that attitude tests (observations of verbal behavior) may be used to predict test performance in English. As a final note, the possibility that Specified Comments may have little positive effect on either test performance or attitudes should serve as a warning to teachers who follow the rather uncreative and depersonalized practice of using "stock" comments, thinking that they are enough. Instead, the teacher should, when time permits, make truly personal comments, comments from which a student can get the feeling that the teacher really is conscious of his efforts, or the lack of them.

CONCLUDING STATEMENTS

A word of caution is advised, however, before overemphasizing the positive results related to the effects of Free Comments. Adding to the fact that the results were not overly significant is the more important consideration of defining "long term" effects. The Page study ran for about two weeks, the present study six weeks. This is quite a few weeks short of a semester, or school year, and any further investigations of the comment variable should extend for a longer period. Future studies of this nature should not have to hide behind an operational definition of "long term." Once it has been decided to execute such a longitudinal study, there is one other important factor which must be investigated. Page made a detailed analysis of factors such as letter grade, school year, and school and found no significant effects due to these factors. However, he failed to control for the sex of both the students and teachers. The present study was blocked controlling for sex, but the analysis of its effects was not within the temporal or statistical scope of this paper. However, study of the ranked data indicated that boys were more affected by the comments than were girls. These data seem to go against both common sense and research evidence. The fact that all three teachers were young females may have had something to do with this. In the future, a longitudinal study controlling for sex will be necessary if a more accurate appraisal of comment effects is desired.

REFERENCES

Aiken, L. R., & Dreger, R. M. The effect of attitudes on performance in math. Journal of Educational Psychology, 1961, 52, 19-24.

Bilodeau, E. A., & Bilodeau, I. McD. Motor skills learning. Annual Review of Psychology, 1961, 12, 243-280.

Bostrum, R. S., Vlandis, J. W., & Rosenbaum, M. E. Grades as reinforcing contingencies and attitude change. Journal of Educational Psychology, 1961, 52, 112-115.

Cohen, A. R. Attitude change and social influence. New York: Basic Books, 1964.

Frankel, E. Studies in teacher-student relationships. Journal of Educational Research, 1960, 53, 172-180.

Friedman, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of The American Statistical Association, 1937, 35, 675-701.

Furness, E. Psychological determinants of spelling success. Education, 1958, 79, 234-239.

Hively, W. Implications for the classroom of B. F. Skinner's analysis of behavior. Harvard Educational Review, 1959, 29, 37-42.

Kennedy, W. A., Turner, A. J., & Lindner, R. Effectiveness of praise and blame as a function of intelligence. Perceptual Motor Skills, 1962, 15, 143-149.

Keuls, M. The use of the "studentized range" in connection with an analysis of variance. Euphytica, 1952, 1, 112-122.

Lindquist, E. F. Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin, 1953.

Lintz, L. M., & Brackbill, Y. Effects of reinforcement delay during learning on the retention of verbal material in adults. Journal of Experimental Psychology, 1966, 71, 194-199.

McDavid, J. Some relationships between social reinforcement and scholastic achievement. Journal of Consulting Psychology, 1959, 23, 151-154.

Miller, N. E. Liberalization of basic S-R concepts. In S. Koch (Ed.), Psychology: A Study of A Science, Vol. II. New York: McGraw-Hill, 1959. Pp. 196-286.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. The measurement of meaning. Urbana: University of Illinois Press, 1957.

Page, E. B. Teacher comments and student performance: A 74 classroom experiment in school motivation. Journal of Educational Psychology, 1958, 49, 173-181. (a)

Page, E. B. Educational research: Replicable or generalizable? Phi Delta Kappan, 1958, 39, 302-304. (b)

Prosad, R. G., & Singh, B. K. Some important characteristics of teachers and students. Psychological Studies, 1962, 7(1), 23-31.

Quay, L. C. Academic achievement and attitudes in group perception in sixth graders. Dissertation Abstract, 1959, 19, 3042-3043.

Schvaneveldt, R. W. Informative feedback and human performance. Madison, Wis.: Author, 1965.

Seashore, H. G., Kurtz, A. K., Kendler, H., Stuntz, S. E., & Rappaport, C. Variation of activities in code classes: An experimental study of the problem of monotony in code learning. OSRD Report 4082, The Psychological Corporation, 1944 (PBL 12173). Cited by D. Wolfle in S. S. Stevens (Ed.), Handbook of experimental psychology. New York: Wiley, 1951. P. 1273.

Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

Skinner, B. F., & Rogers, C. R. Some issues concerning the control of human behavior. Science, 1956, 124, 1057-1066.

Skinner, B. F., & Holland, J. G. The use of teaching machines in college instruction. In A. A. Lumsdaine and R. Glaser (Eds.), Teaching machines and programmed learning, Washington, D. C.: Dept. of Audio Visual Instruction, NEA, 1960.

Solomon, D., & Rosenberg, L. Teacher-student feedback and classroom social structure. Journal of Social Psychology, 1964, 62, 197-210.

Staats, A. W., & Staats, C. K. A comparison of the development of speech and reading behaviors with implications for research. Child Development, 1962, 33, 831-846.

Staats, A. W., & Staats, C. K. Complex Human Behavior. New York: Holt, Rinehart and Winston, 1963.

Sweet, R. C. Educational attainment and attitudes toward school as a function of feedback in the form of teachers' written comments. Unpublished master's thesis, University of Wisconsin, 1966.

Weaver, C. H. Semantic distance between students and teachers and its effects upon learning. Speech Monograph, 1959, 26, 273-281.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Wolfle, D. Training. In S. S. Stevens, (Ed.), Handbook of Experimental Psychology. New York: Wiley, 1951. Pp. 1267-1286.

Wright, W. W., & Jung, C. W. Why capable high school students do not continue their schooling. Bulletin of School Education, Indiana University, 1959, 35, 1-78.